

# E-VChat : 頭部動作連動型音声駆動身体引き込みキャラクタを 対面合成した実映像対話システム

石井 裕<sup>\*1</sup> 高田 友寛<sup>\*2</sup> 渡辺 富夫<sup>\*1 \*3</sup>

E-VChat: A Video Communication System in Which a Speech-driven Embodied Entrainment Character Working with Head Motion is Superimposed on the Other Talker's image in a Virtual Face-to-Face Scene

Yutaka Ishii<sup>\*1</sup>, Tomohiro Takada<sup>\*2</sup>, and Tomio Watanabe<sup>\*1 \*3</sup>

**Abstract** – We proposed an embodied video communication system in which a human-type avatar called “VirtualActor” which represents interactive behavior is superimposed on the other speech partner's video image in a virtual face-to-face scene. The effectiveness of a video communication system was demonstrated in an experiment of comparison with the scene in which a reduced own video image is superimposed on the other talker's video image using the picture-in-picture method. However, this system had some problems, such as the detailed adjustment of video images and the lack of portability of sensors. In this paper, we develop a headset-type motion-capture device which reflects the talker's head movements directly using an acceleration sensor and gyro sensor, and employ a CG character which moves based on talker's own motion and generates motion automatically based on the on-off pattern of talker's voice. Further, we propose the concept of an embodied video communication system in which the CG character is superimposed on the other talker's video image in a face-to-face scene, and develop a prototype called “E-VChat”. A communication experiment is performed to confirm the effectiveness of the E-VChat system for 12 pairs of subjects using three communication modes: “Headset,” “Headset + Generated motion automatically as a talker's avatar,” and “Headset + Generated motion automatically as a talker's support agent.” The results show that all communication modes tested are affirmatively assessed by sensory evaluation, and the “Headset + Generated motion automatically as a talker's avatar” mode is evaluated highly by a paired comparison. Finally, we develop a multiple-character E-VChat system using an audience that nods in response to the talker's voice, and confirm the effectiveness of the system in an interview-style communication experiment.

**Keywords** : Embodied Interaction, Video Communication, Avatar Mediated Communication, Head Motion Tracking, Voice-driven Embodied Character

## 1. はじめに

年々向上する情報通信インフラは、動画コンテンツなど大容量データを高速に送受信することを可能とし、遠隔での多様なコミュニケーションを創出している。携帯電話についても音声通話のみならず Web を介した動画閲覧など高速なネットワークが活用され、特にこれらの取り扱いに特化した多機能携帯電話（スマートフォン）の利用者も増加し、リッチコンテンツによるデータ通信量も増加し続けている<sup>[1]</sup>。

コミュニケーションメディアとしてのビデオ映像は、遠隔コミュニケーションにおいて対話者のうなずきや、

身振り・手振り、表情といった言葉によらないノンバーバル情報を直接的に把握できるなど、大いに有用といえる<sup>[2],[3]</sup>。ビデオ対話システムは友人や家族間などで利用が拡大しつつあり、またビデオ映像を利用した遠隔ビデオコミュニケーションとしても多くのテレビ会議システムが提案され、様々な会議や講義等の形態に応じたシステムの研究開発や利用効果の検討がなされている<sup>[4]~[8]</sup>。

一方で、新しいメディア表現としての映像効果の開発は、その魅力を改めて引き出すことにつながる。ここでこの映像を用いたコミュニケーションにおいて、従来のように相手のみの映像あるいは自己映像を P-in-P (Picture-in-Picture) で挿入するだけでなく、対話感を重視した効果的な画面構成とすることで、インタラクションを促進し革新的な対話メディアを実現することで新たな利用場面を創出することができると考えられる。

\*1: 岡山県立大学 情報工学部

\*2: 岡山県立大学大学院 情報系工学研究科

\*3: 科学技術振興機構 戦略的創造研究推進事業

\*1: Faculty of Computer Science and Systems Engineering, Okayama Prefectural University

\*2: Graduate School of Systems Engineering, Okayama Prefectural University

\*3: CREST of Japan Science and Technology Agency

効果的なインタラクション支援を目的として、著者はこれまでに対話相手のビデオ映像に対して、自己映像を色相編集技術であるクロマキー手法を用いて仮想対面合成したビデオコミュニケーション手法および対話者の身体動作を忠実に再現するアバタとしてのCGキャラクタである VirtualActor を重畳合成する身体的ビデオコミュニケーションシステムを開発してきた<sup>[9]</sup>。但しこの VirtualActor は磁気センサを装着して実現する対話者自身のアバタであり、動作を忠実に再現することが可能であるものの、拘束感や画面上の違和感、また可搬性に大きな問題を抱えている。またモバイル端末への実装を検討した場合、身体動作の計測はかなり限定的になる可能性が高い。

そこで本論文では、動作計測部位をインタラクションに重要な役割を果たす頭部のみに限定して機器の簡略化と装着感の向上を図り<sup>[10]</sup>、身体全体の動作は音声と動作の相互相関関係に基づく自動生成モデルによって構成することにより、インタラクションのリズムを補うCGキャラクタを用いた実映像対話システムを構築した。CGキャラクタに対話相手の映像に重畳合成することにより、アバタとしての意図的な動作を含むインタラクティブなコミュニケーション動作の効果的な利用により新たな映像表現としてのビデオチャットシステムを構成している。また12組24人を対象としたコミュニケーション実験においてコミュニケーション支援の観点から官能評価を行い、開発したシステムの有効性を示している。さらに緊張した対話場面での利用を目的として、初対面同士の対話に対して観客キャラクタを含む複数のCGキャラクタを配置したインタビュー実験を行い、そのコミュニケーション効果を検討している。

## 2. ビデオコミュニケーションにおける自己像の重畳合成

従来より二者間でのコミュニケーションにおいては、映像を用いたコミュニケーションモードとして相手のみの映像あるいは自己映像をP-in-Pで挿入された画面が利用されているが、これは実環境での対面対話のような同じ空間での対話と異なり、分離された画面構成では互いのインタラクションを把握しにくいという問題がある<sup>[12], [13]</sup>。ビデオコミュニケーションにおけるこれらの問題を解決するために、対話者同士を同一空間に配置するために自己映像を重畳合成したシステム開発が提案されている<sup>[14], [15]</sup>。

著者らもこれまでに、対話者間の円滑なインタラクションを支援するために、図1に示す対話相手及び対話者自身をビデオカメラで撮影した映像を仮想的に対面合成したビデオコミュニケーション手法を提案し、



図1 斜め後方からの自己映像を重畳合成したビデオコミュニケーション手法

Fig.1 Video communication method in which own video image is superimposed on the other talker's video image from the diagonal view-point.



図2 VirtualActorを重畳合成した身体的ビデオコミュニケーションシステム

Fig.2 Embodied video communication system in which own VirtualActor is superimposed on the other talker's video image.

対話相手のみの正面映像と比較して手法の有効性を確認している<sup>[9]</sup>。しかし対話者の詳細な映像を利用することは、対話者間のサイズ、明度・輝度差などの映像情報に対して緻密な調整が必要となる。細かなずれが画面に対して違和感を生じ、インタラクションに影響を与える結果を生む。そのため、対話者自身の代役として、図2に示すように対話者の身体動作を忠実に再現するアバタとしてのCGキャラクタである VirtualActor と、対話相手をビデオカメラで撮影した映像を色相編集技術であるクロマキー手法を用いて仮想対面合成した身体的ビデオコミュニケーションシステムを開発し、コミュニケーション実験において官能評価及び行動分析によりシステムの有効性を示した<sup>[9]</sup>。これは映像の詳細な調整をすることなく互いの身体的インタラクションが捉えやすい状態を実現するため、ビデオ映像を用いて対話相手の表情や身体動作など、ノンバーバル情報を観察しながらのコミュニケーションを可能にするシステムである。

しかしこのシステムでは対話者の頭部、腰部、両手首の4箇所に磁気センサ (Polhemus FASTRAK) を装

着し、人の上半身の動作を忠実に再現するキャラクタ重畳合成システムであり、センサの装着による拘束感や画面上の違和感、可搬性に大きな問題がある。このため、VirtualActorの動作をユーザの音声と頭部動作に基づいて自動生成することで、実映像を使用しつつ手軽で快適なパフォーマンスを実現する、相手映像に自己の代役となるキャラクタを重畳合成する Enhanced VideoChat(E-VChat) システムを開発している。次章でその詳細を述べる。

### 3. E-VChat システム

#### 3.1 コンセプト

本研究で開発した Enhanced VideoChat(E-VChat) システムのコンセプトを図3に示す。対話相手をカメラで撮影した映像に対して、対話者自身の代役キャラクタを重畳合成したシステムである。画面上の構成はこれまでに著者らが提案した VirtualActor を重畳合成した身体的ビデオコミュニケーションシステムと同様である。

E-VChat システムでは、拘束感や画面上の違和感、可搬性に対する問題を解決するために、代役のキャラクタ動作の半自動化により身体動作を補完的に表現する。装着時に違和感のない小型・軽量のセンサ装着によって対話者自身の頭部動作を忠実に再現するとともに、マイクから入力された音声を二値化した ON-OFF データの時系列からうなずき、瞬き、身体動作のタイミングを推定する MA(Moving-Average) モデルに基づいて生成されたコミュニケーション動作を行うキャラクタを自動生成し、対話リズムをよりインタラクティブに支援するシステム構成とする<sup>[16]</sup>。交渉や協議など細心の注意が必要な即応性の高い対話や、内容の信頼性が重視されるキャラクタを介した情報発信などの用途においては、自動生成されたコミュニケーション動作だけでなく、対話者の意思を直接的に示す必要がある場合がある。そのためコミュニケーションに重要な役割を果たす頭部動作をキャラクタに連動させ、肯定・否定の意思表現を反映可能にすることで、よりキャラクタを介したコミュニケーション特性を生かしたシステム構築を行うことができると考えられる。

ここでキャラクタ重畳合成の構成として、図4に示すようにカメラをモニタ上部に置くことで、図5のようにカメラは自然とモニタを見る対話者を見下ろす角度となる。従来は大きな問題となっていた対話相手の視線のずれを利用し、画面の下部中央に配置することで、正面から撮影される対話相手が自己のキャラクタと対話しているように見える。

本システムの特徴は、キャラクタを観察するのはあくまで対話者自身だけであることにある。画面上対話

相手が自己の代役キャラクタを観察しているように見えるが、実際には対話相手も対話相手自身の代役キャラクタを観察しており、キャラクタの振る舞いは対話相手に伝達されない。そのためキャラクタによる効果は直接的には対話者自身にのみ起こりうるが、結果的にその効果は対話者相互のインタラクションに影響を与えると考えられる。本システムは対話者自身に対して効果的な映像表現を提示することにより、その特徴的な効果を利用してコミュニケーションを支援する。

#### 3.2 音声に基づく自動生成によるキャラクタ動作モデル

まずキャラクタの音声に基づくコミュニケーション動作の自動生成モデルは、これまでに著者らが提案してきた音声の ON-OFF の時系列からうなずき、瞬き、身体動作のタイミングを推定する MA モデルを使用する<sup>[16]</sup>。音声データは 16bit 22.050kHz でサンプリングし、閾値で二値化するとともに、音節間の短時間の無音区間による発話の断片化を除去するために 133msec でハングオーバー処理を施している。

この音声に基づくコミュニケーション動作は「話し手」及び「聞き手」の動作モデルによって生成される。まず「聞き手」のインタラクションモデルとして音声の ON-OFF パターンに基づくうなずき反応モデルを導入している(図6)。このモデルは、マクロ層とミクロ層からなる階層モデルを用いてうなずきの予測を行う。マクロ層では音声の呼気段落区分での ON-OFF 区間からなるユニット区間にうなずきの開始が存在す

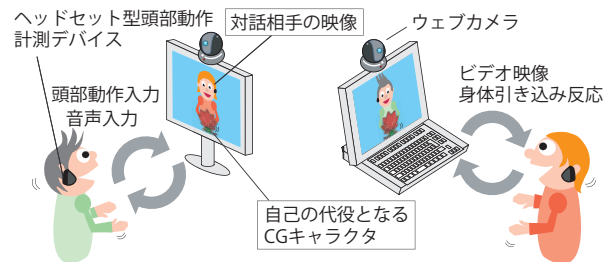


図3 E-VChat システムのコンセプト  
Fig. 3 Concept of E-VChat system.

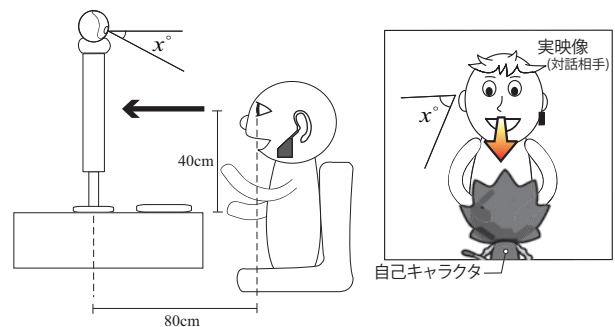


図4 対話者の視線位置模式図  
Fig. 4 Talker's line of vision.

図5 画面上での目線  
Fig. 5 Line of vision on the screen.

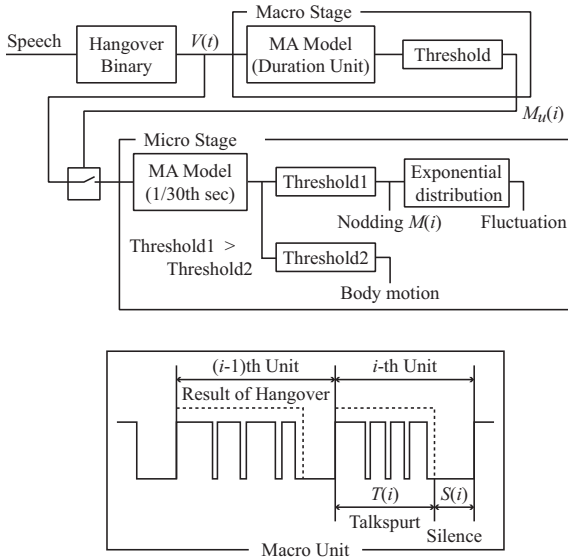


図6 うなずき反応モデル  
Fig. 6 Nodding response model.

るかを  $[i-1]$  ユニット以前のユニット時間率  $R(i)$  (ユニット区間での ON 区間の占める割合, (1) 式) の線形結合で表される (2) 式の MA(Moving-Average) モデルを用いて予測する. 予測値  $M_u(i)$  がある閾値を越えて, うなずきが存在すると予測された場合には, 処理はミクロ層に移る. ミクロ層では音声の ON-OFF データ (30Hz, 60 個) を入力とし, (3 式) を用いて MA モデルでうなずきの開始時点を推定する. 予測値が閾値を越えた場合にキャラクタがうなずき動作を行う. 身体動作についてもこの予測値に基づき, うなずきよりも低い閾値で各部位 (頭部, 胴部, 右肘, 左肘) のうちいずれかを選択して動作させることでうなずきと関連付けている.

$$M_u(i) = \sum_{j=1}^J a(j) R(i-j) + u(i) \quad (1)$$

$$R(i) = \frac{T(i)}{T(i) + S(i)} \quad (2)$$

$a(j)$ : 予測係数  
 $T(i)$ :  $i$  番目ユニットでの ON 区間  
 $S(i)$ :  $i$  番目ユニットでの OFF 区間  
 $u(i)$ : 雑音

$$M(i) = \sum_{j=1}^K b(j) V(i-j) + w(i) \quad (3)$$

$b(j)$ : 予測係数  
 $V(i)$ : 音声データ  
 $w(i)$ : 雑音

また「話し手」は, 対面での自由対話を計測して対話者自身の発話音声の ON-OFF パターンに基づいて, 入力音声に対する頭部動作および腕や胴体などを動作させることで発話音声と関連付けた. さらに, 対話者自身の音声に対して聞き手動作としてフィードバックさせることも可能であり, 対話を促進するエージェント的役割として自己の発話音声に対する応答・共感反応を行うキャラクタとして存在させることができる. これらの効果的な対話リズムの生成により, 対話者自身に会話意欲の促進を働きかけ, コミュニケーション場の形成を支援する.

### 3.3 ヘッドセット型頭部動作計測デバイスによる頭部動作合成

3.1 節で述べた通り, 自動生成されたコミュニケーション動作だけでなく対話者の意思を直接的に示す必要がある場合, ビデオ映像に違和感が無いように目立たない頭部動作計測デバイスの開発が必要となる. 著者らはこれまでに, 耳にフックして使用するヘッドセット型動作計測デバイスを開発している<sup>[10]</sup>. このデバイスを用いて頭部動作を計測し, 3.2 節で述べた MA モデルによる自動生成された動作に対して重畳合成する. 装置の外観と内部構成を図 7(a) に示す. センサは 3 軸加速度センサモジュール (秋月電子通商 KXM52-1050) とジャイロセンサユニット (近藤科学 KRG-3) であり, それぞれ頭部のうなずき方向と頭部の横振り方向を計測する. うなずき方向の計測は頭部動作の特性に基づき加速度センサで頭部の傾きを検出し, 横振り方向の頭部動作はジャイロセンサによる頭部の回転角速度の検出によって算出している.

接続部であるインタフェース装置の外観と内部を図 7(b) に示す. センサからの信号をマイコンモジュール (秋月電子通商 USB マイコンボード) により A/D 変換し PC へ入力している. インタフェース装置にはクリップが付いており, 使用者の衣服に固定することができる. ケーブルを除いたヘッドセットとセンサ部は, 縦 102mm, 横 61mm で, 重量は約 19g である.

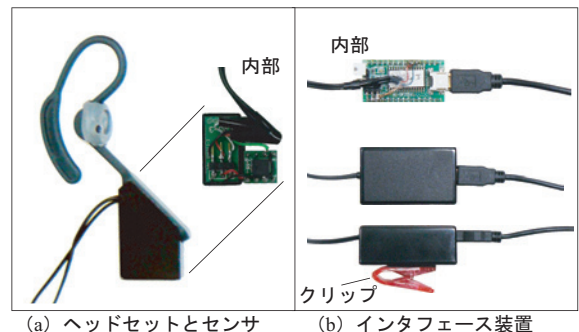


図7 ヘッドセット型頭部動作計測デバイス  
Fig. 7 Headset-type motion capture device.



この計測装置を用いてキャラクタのみを介した対話において検討を行った結果、「楽しさ」「キャラクタになった感覚」などの項目で高く評価されており、装着時における違和感で会話を阻害する影響が少ないことが確認されている<sup>[10]</sup>。

### 3.4 プロトタイプシステム構成

開発したシステムのプロトタイプを図8に示す。対話者にとって中央に、対話相手の上半身と、自己の代役となるキャラクタを表示させた。キャラクタによる動作特性については、よりプリミティブなオブジェクトとしての円柱・円盤型キャラクタにおいて同様の効果を確認しており<sup>[11]</sup>、これまでに様々なキャラクタを用いてシステム開発を行ってきた。本研究ではキャラクタのデザインとして、多様なコミュニケーション場面に応じた効果的な画面構成を演出する目的で、人型に限らず抽象化された動物を利用することとし、予備検討で多く好まれた2.5頭身のライオン型を選択している。キャラクタは対話者の代役とするため、垂直軸方向に180度回転させて仮想的に相手と正面から対面させた。



図8 プロトタイプシステム画面

Fig.8 Screen shot of the prototype system.

## 4. システム評価実験

### 4.1 コミュニケーション実験概要

開発したシステム評価を目的として、システムを使用して別々の個室に分かれて日常会話を行うコミュニケーション実験を行った。実験環境の全体構成を図9に示す。実験に使用したPCはFujitsu FMV LIFEBOOK NH900(CPU: Intel Core i7 2.80GHz, Memory: 4GB, Graphic Board: GeForce GT 335M, Windows 7 Professional 32bit)であり、CGキャラクタの描画フレームレートは30f/sである。対話者間は1GB/sのイーサネットに接続されており、音声通信を行った。また、対話者が互いに映像を送るため、USB接続されたウェブカメラ(Logicool QcamR Orbit AF QCAM-200RX)を使用し、PCモニタの上部中央に設置した。実験の様子は被験者後方からビデオカメラで撮影した(図10)。

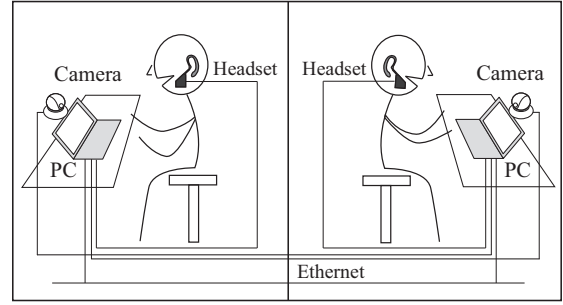


図9 実験概略図

Fig.9 Experimental setup.



図10 コミュニケーション実験場面

Fig.10 Communication scene in the experiment.

表1 実験で使用した3モード

Table 1 3 communication modes in the experiment.

A	ヘッドセット反映のみ
B	ヘッドセット+動作自動生成 (自己音声に対する話し手+相手音声に対する聞き手)
C	ヘッドセット+動作自動生成 (自己音声に対する聞き手)

評価実験において、表1に示す重畳合成する自己キャラクタの動作が異なる3種類のモードを設定した。それぞれ、A：ヘッドセットによる対話者の頭部動作を反映するモード、B：ヘッドセットに加えて対話音声に基づき自己の発話音声に対する話し手動作と相手発話音声に対する聞き手動作を自動生成して合成するモード、C：ヘッドセットに加えて対話音声に基づき自己の発話音声に対する聞き手動作をするモードである。Aは従来開発してきたVirtualActorによるシステムを、小型化、軽量化したヘッドセット型動作計測装置により改良したモードであり、B、Cは発話音声に基づいて自動生成された動作を補完的に合成した構成となっている。Bは対話者のアバタ的役割を行う自己の発話音声に対する話し手動作、相手の発話音声に対する聞き手動作であり、Cは視覚的に対話相手とのインタラクションを支援する配置にありながら自己の発話音声に聞き手として動作して対話リズムを支援する

コミュニケーションエージェント的役割を行う。このモードCについては、画面内の対話相手と対面合成する配置を取った状態で、自己の音声に対してより発話を促す効果のあるうなずき動作を自動生成した場合のコミュニケーション効果を検証する目的で設定した。

被験者は従来システムの使用経験の無い20～24歳の男女学生で、同性同士の友人関係である12組24人である。対話内容は、システム評価に関する話題を除き、制約を設けない自由対話とした。

実験前に各モードの違いと操作方法を説明し、システムを使用させながら確認させた。その後20秒程度ずつ実際に対話させながらモードの切替に関する練習時間を設けた。モードの切替等実験中の指示は、実験観察者が室外からマイクとスピーカで行った。

実験はまず、各モードからペア毎にカウンターバランスを取った2つのモードを抽出し、約1分間ずつ対話させた後、総合的な判断で一対比較によって良かったモードを選択させた。次に、各モードに対し1つのモード毎に約1分間ずつ対話させた後、身体的コミュニケーション支援の観点から定めた5項目「楽しさ」「一体感」「対話しやすさ」「安心感」「好み」について7段階評価(+3～-3, 中立0)させた。

#### 4.2 官能評価結果

一対比較評価の結果を表2に示す。表中の数字は各モードで選択された回数を表している。この結果を定量的に評価するためにBradley-Terryモデルを想定し(式4)、強さ $\pi$ を最尤推定した結果を図11に示す。

表2 一対比較結果  
Table 2 Result of pair comparison.

	A	B	C	計
A		4	7	11
B	20		20	40
C	17	4		21

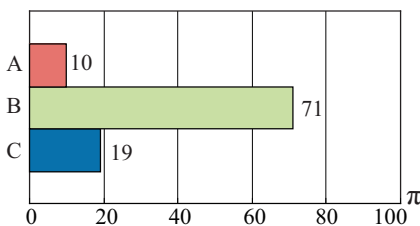


図11 Bradley-Terryモデルによる好ましき評価  
Fig.11 Preference based on the Bradley-Terry model.

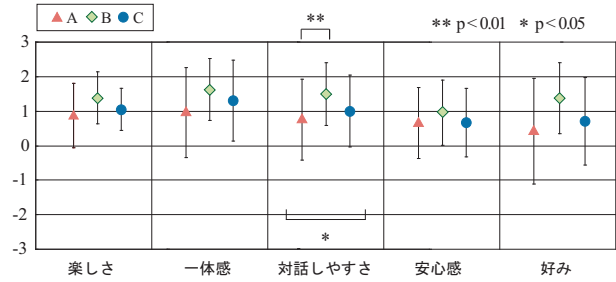


図12 7段階評価  
Fig.12 Seven points bipolar rating.

$$P_{ij} = \frac{\pi_i}{(\pi_i + \pi_j)} \quad (4)$$

$$\sum_i \pi_i = \text{const.} (= 100)$$

ここで $\pi_i$  :  $i$ の強さ,  $P_{ij}$  :  $i$ が $j$ に勝つ確率である。

モデルを適合度検定により確認した結果,  $P = 0.32 > 0.05$ でモデルは棄却されず, 有意性検定において $P = 1.67 \times 10^{-6} < 0.05$ で「対象間に差が無い」とする仮説を棄却した。よって図に示す通り, Bの「ヘッドセット+動作自動生成(自己音声に対する話し手+相手音声に対する聞き手)」が最も高く評価されており, 続いてC, Aの順で評価されている。

また, 楽しさ, 一体感, 対話しやすさ, 安心感, 好みの5項目に対する7段階評価の結果を図12に示す。5項目全てにおいてどのモードも平均的に高く評価されており, 特にモードBが高く評価されていることがわかる。Friedmanの分散分析による検定の結果, 「対話しやすさ」の項目で有意水準5%で有意差が認められ, さらに多重比較としてのWilcoxonの符号付順位和検定により, モードAとBの間に有意水準1%で有意差が認められた。

#### 4.3 考察

官能評価結果より, 3モード全て5項目で肯定的に評価されていることから, 著者らが従来より提案する自己アバタを仮想的に対面合成するビデオコミュニケーション手法の有効性が再確認された。また特に本論文で提案する自己音声に対する話し手と相手音声に対する聞き手として動作するキャラクタによるシステムが高く評価されている。「対話しやすさ」の項目のみでモードAとB間に有意差が確認され, ヘッドセット型計測装置によって意識的な動作が可能であることが対話しやすさ評価に反映されたものと考えられる。他の項目については実映像とCGキャラクタの合成場面比較に対して大きな差異が確認されなかった。

ここでアンケート用紙に設けた自由記述欄の回答に

おいても、「自分の動きに合わせてキャラクターが動いてくれるので話しやすかった」「ライオンの手まで動いているのが良かった」といったシステムを有効に活用した意見が確認された。

また「ライオンより人に目がいってしまうけれど、自然にライオンが動いてくれると話しやすかった」といった対話相手への注視を重視した意見や仲介役としてのキャラクタ効果に対するコメントの他、「ライオンの表情も見てみたい」という後ろ向きのキャラクタに対する要望や、「Cのモードの時（自己音声に対する聞き手動作）、ライオンがこちらを向いてくれると良いと思った」というキャラクタ配置よりも動作特性による効果を優先するコメントも見られた。このように対話相手のビデオ映像へのキャラクタの重畳合成は、その利用状況に合わせた特性適用により様々な効果をもたらす可能性があると考えられる。

## 5. 複数キャラクタ配置による対話支援

### 5.1 システム概要

ビデオ映像を用いたコミュニケーション場面としては、前章で想定した友人同士の自由対話だけでなく、初対面の人との対話など緊張した対話場面での利用が想定される。そこで自己アバタを単体で使用するだけでなく、緊張緩和のために対話支援キャラクタを複数体配置するシステムとして応用を検討する。キャラクタを複数体配置することで、実映像と仮想的に重畳合成して集団によるコミュニケーション場を形成し、場の盛り上げなどのコミュニケーション効果を応用したコミュニケーション支援が期待できる。

前章で構築した自己アバタとしての頭部動作連動型音声駆動身体引き込みキャラクタに加えて、対話者の音声にのみ反応する4体の音声駆動身体引き込みキャラクタを配置したシステムを図13に示す。自己の代役キャラクタと明確に区別するため、別のキャラクタデザインとした。ここではより緊張感のある対話を想定して、緊張感を緩和するために捕食-被食関係にあるライオンと羊をモデルとして対話相手に対する優位性の演出効果を期待するデザイン構成とした。言い換えれば、観客キャラクタに対する優位性を自己アバタに与えるデザインとすることで、より積極的な発話が支援できる可能性がある。対話支援としての羊キャラクタの注視方向は、対話者自身に向ける、いわゆるカメラ目線ではなく、代役キャラクタであるライオンに向けられている。これは対話相手の映像を含めた画面上のインタラクション場を整合的に表示するためである。キャラクタ配置については、相手映像と対面合成して違和感のないライオンのサイズを基準に、同程度のサイズで相手映像の隠蔽が少なくバランスの良い4



図13 複数体配置システム  
Fig. 13 Multiple-character system.

体の配置とした。

### 5.2 インタビューによるコミュニケーション実験

キャラクタ複数体配置システムの評価を目的として、インタビュー形式によるコミュニケーション実験を行った。システムを正確に評価させるために、被験者に負荷をかけ過ぎない程度の状況を設定するため、一対一のインタビュー形式とした。初対面の上級生と下級生という組み合わせで、上級生を質問者、下級生を回答者として2学年以上の差となるように設定した。被験者は同性同士の18~24歳の男女学生12組24人である。被験者には、実験の主旨を「充実した学生生活を送るために上級生から下級生へインタビューすることと説明を行い、質問者としての上級生には予め参考となる質問例（「休日はどのように過ごしていますか」「どのような職業に就きたいですか」など）を渡すとともに、回答者の返答によって適宜アドバイスを与えるように指示した。経験によるこのアドバイスをより現実的・効果的に返答させるために、2学年以上の差を設定している。

実験には図14に示す3モードを使用した。それぞれ、I: 対話相手のみ、II: 自己アバタ単体の対面合成、III: 今回開発した自己アバタおよび4体の観客キャラクタを重畳合成、という構成である。前章での実験同様、実験開始前に約15分程度でシステム説明及びテスト使用を行ってシステムに慣れさせた後、総合的な判断で一対比較によって良かったモードを選択させた。次に、各モードに対し約3分間ずつ対話させた後、「対



図14 インタビュー実験に使用した3モード  
Fig. 14 3 communication modes used in the interview experiment.



話しやすさ」「一体感」「安心感」「好み」「満足感」の5項目と、質問者には「話を引き出せたか」、回答者には「主張しやすさ」の項目を加えた6項目について7段階評価(+3~-3, 中立0)させた。

### 5.3 官能評価結果

一対比較評価の結果を表3に示す。また、式4のBradley-Terryモデルにより強さ $\pi$ を最ゆう推定した結果を図15に示す。モデルを適合度検定により確認した結果、 $P = 0.67 > 0.05$ でモデルは棄却されず、有意性検定において $P = 1.65 \times 10^{-5} < 0.05$ で「対象間に差が無い」とする仮説を棄却した。よって図に示す通り、モードIIIの「自己アバタおよび4体の観客キャラクタを重畳合成」が最も高く評価されており、続いてII, Iの順で評価されている。

次に6項目に対する7段階評価の結果について、質問者の結果を図16に、回答者の結果を図17に示す。質問者について、Friedmanの分散分析による検定の結果、「対話しやすさ」「安心感」「好み」「満足感」の項目で有意水準1%で有意差が認められた。多重比較としてのWilcoxonの符号付順位和検定により、「対話しやすさ」でモードIとIIIの間に有意水準1%で有意差が認められ、「安心感」ではモードIとIIIの間に有意水準1%で、モードIとII, IIとIIIの間で有意水準5%で、「好み」「満足感」ではモードIとIIIの間で有意水準5%で有意差が認められた。今回開発したモードIIIが全体的に肯定的に評価され、モードIの相手のみの映像は否定的に評価されていることが分かる。

また回答者については、Friedmanの分散分析による検定の結果、「相手との一体感」「安心感」の項目で

表3 インタビュー実験における一対比較結果  
Table 3 Result of pair comparison in the interview experiment.

	I	II	III	計
I		7	3	10
II	17		8	25
III	21	16		37

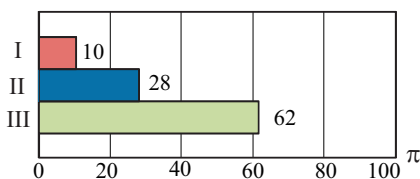


図15 インタビュー実験における Bradley-Terry モデルによる好ましき評価

Fig.15 Preference based on the Bradley-Terry model in the interview experiment.

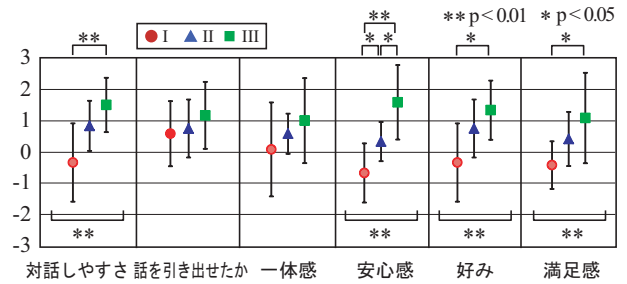


図16 質問者側の7段階評価

Fig.16 Seven points bipolar rating for interviewers.

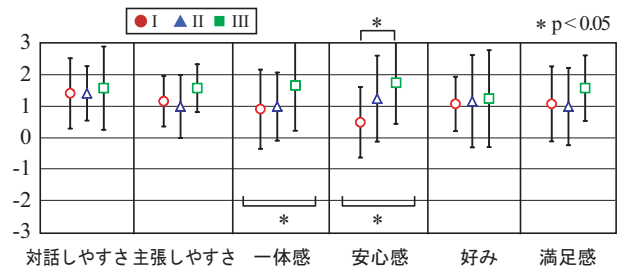


図17 回答者側の7段階評価

Fig.17 Seven points bipolar rating for interviewees.

有意水準5%で有意差が認められ、多重比較としてのWilcoxonの符号付順位和検定により、「安心感」の項目でモードIとIIIの間に有意水準5%で有意差が認められた。回答者はモードIも含めて全てのモードが肯定的に評価されている。

### 5.4 考察

本実験では、質問者に対して「返答によって下級生にアドバイスを与える」ことをタスクとして設定した。それにより質問者は自己のアドバイスとして相手の持っていない情報、的確なコメントを提供する必要があり、回答者よりも対話内容を強く意識して緊張を伴った可能性がある。ここで今回開発したモードIIIは、自己音声に対して効果的にうなずき動作を行う複数のキャラクタが周囲に配置されたデザインであり、質問者は自己の発言に対する周囲の羊キャラクタの応答を肯定的な反応として受け止めたため「安心感」が得られ、発話を支援するシステムとして高く評価したと考えられる。つまり、上級生と下級生という社会的な立場における緊張緩和としてのインタラクションデザイン効果について、本実験における質問者側の評価で「対話しやすさ」「安心感」「好み」「満足感」の項目で有意差が確認されたことで、質問者役に不慣れな被験者にとくに効果的であったと言える。これは「話を引き出せたか」「一体感」の項目で差が確認されなかったことから、質問者として話を引き出すことよりも、自己のアドバイスをうまく提示できるかどうかを意識している傾向がうかがえる。ここで自由記述欄に



おいても「キャラクタがあることで初対面の人とも和んで話すことができた」「Cは羊に囲まれていて安心感を感じる」などの回答が得られ、キャラクタを介することによる緊張緩和効果が確認された。

また「Iのビデオチャットだけでは不安な気がした」「Iのとき目線が下なのが話していて（聞いていて）気になる。II, IIIは相手に集中しているから変わった印象を受けなかった」など、自己アバタによりビデオ映像に対する視線のずれを補完する配置を評価するコメントも確認された。

複数のキャラクタを使用した集団としての効果は様々な面で応用可能であり、対話者自身の音声のみ、あるいは対話相手の音声のみに反応させるなど、対話場の表現としても効果的な利用が可能である。しかし本実験ではインタラクションの観点から緊張緩和に着目して分析評価を行ったため、集団コミュニケーションとしてのキャラクタデザインの影響については更に詳細な分析が必要である。これまでに教育支援を目的として、教育番組などの映像に聴講者としてのCGキャラクタを重畳合成し、最適なキャラクタの数、配置、大きさについての検討なども行われており<sup>[17]</sup>、今後映像に対するキャラクタ合成による効果を体系的に分析する。さらに本論文では自由対話及びインタビュー形式により実験を行ったが、キャラクタ動作の自動生成モデルについても、相手を説得するディベート形式など、対象とするコミュニケーション場面に応じて最適な反応、動作量等のパラメータを検討する必要がある。

## 6. おわりに

本論文では、動作計測部位をインタラクションに重要な役割を果たす頭部のみに限定して機器の簡略化と装着感の向上を図るとともに、身体全体の動作は音声と動作の相互相関関係に基づく自動生成モデルによって構成することにより、インタラクションリズムを効果的に利用するCGキャラクタを用いたビデオチャットシステムE-VChatを提案した。対話相手の音声に対して聞き手として反応し、対話者自身の音声に対して話し手として反応するアバタ的動作、及び対話者自身の音声に対して聞き手として反応するエージェント的動作を行うシステムを実装し、12組24人を対象としたコミュニケーション実験により、本システムの有効性を示した。また複数キャラクタを利用して集団コミュニケーション場を構成する複数体重畳合成システムを開発し、緊張した対話場面としてのインタビュー形式によるコミュニケーション実験における官能評価により有効性を示した。

## 謝辞

本研究は科学技術振興機構（JST）の戦略的創造研究推進事業（CREST）研究領域「デジタルメディア作品の制作を支援する基盤技術」における「人を引き込む身体性メディア場の生成・制御技術」プロジェクトの支援による。

## 参考文献

- [1] 総務省:平成22年度 通信利用動向調査報告書, (2011).
- [2] 黒川 隆夫: ボディ表現とマルチメディア, 自己の表現 (岩波講座マルチメディア情報学10), pp.235-294 (2000).
- [3] 大坊 郁夫: しぐさのコミュニケーション, サイエンス社, pp.16-90 (1998).
- [4] Sellen, A. J.: Speech Patterns In Video-Mediated Conversations, Proc. of CHI' 92, ACM, pp.49-59 (1992).
- [5] William A. S. Buxton: Living in Augmented Reality: Ubiquitous Media and Reactive Environments, Video-Mediated Communication (Edited by Finn, E. K., et al.), Computers, Cognition, and Work, pp.363-384 (1997).
- [6] 井上 智雄, 岡田 謙一, 松下 温: テレビ会議における映像表現の利用とその影響, 情報処理学会論文誌, Vol.40, No.10, pp.3752-3761 (2000).
- [7] Ishii, R., Ozawa, S., Mukouchi, T., and Matsuura, N.: MoPaCo: Pseudo 3D Video Communication System, Human Interface, Part II, HCI 2011, LNCS 6772, pp. 131-140 (2011).
- [8] 磯 和之, 伊達 宗和, 高田 英明, 安藤 康子, 松浦 宣彦: 視線の向きを表現可能な2画面積層表示を用いたテレビ会議システムの提案, 情報処理学会論文誌, Vol.52, No.3, pp.1224-1233 (2011).
- [9] 石井 裕, 渡辺 富夫: VirtualActorを対面合成した身体的ビデオコミュニケーションシステム, ヒューマンインタフェース学会論文誌, Vol.5, No.2, pp.225-234 (2003).
- [10] Yamamoto, M., Osaki, K., Matsune, S., and Watanabe, T.: An Embodied Entrainment Character Cell Phone by Speech and Head Motion Inputs, Proc. of the 19th IEEE International Symposium in Robot and Human Interactive Communication Symposium, pp.318-323 (2010).
- [11] 吉田 真章, 渡辺 富夫, 山本 倫也: 3DCGオブジェクトを用いた音声駆動型身体的引き込みシステム, ヒューマンインタフェース学会論文誌, Vol.9, No.3, pp.369-378 (2007).
- [12] 松尾 太加志: コミュニケーションの心理学, ナカニシヤ出版, pp.71-135 (1999).
- [13] 原田 悦子編著: 「使いやすさ」の認知科学, 共立出版株式会社, pp.55-98 (2003).
- [14] 関 秀寿, 服部 真承, 村田 和義, 渋谷 雄: ミラーインタフェースにおける触覚フィードバックの導入とその遅延がユーザビリティに与える影響, ヒューマンインタフェース学会論文誌 Vol.12, No.2, pp.21-29 (2010).
- [15] 森川 治, 橋本 佐由理, 前迫 孝憲: 仮想的な抱擁を取り入れた遠隔カウンセリングシステム, 日本バーチャルリアリティ学会論文誌, Vol.14, No.1, pp.3-10 (2009).
- [16] Watanabe, T., Okubo, M., Nakashige, M., and

Danbara, R.: InterActor: Speech-Driven Embodied Interactive Actor; International Journal of Human-Computer Interaction, Vol.17, No.1, pp.43-60 (2004).

- [17] 山本 倫也, 渡辺 富夫: 音声駆動型身体引き込みキャラクターを映像に重畳合成した教育支援システム, 情報処理学会論文誌, Vol.47, No.8, pp.2769-2778 (2006).

(2012年3月22日受付, 6月25日再受付)

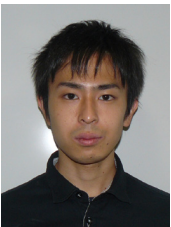
## 著者紹介

### 石井 裕 (正会員)



1998年岡山県立大学情報工学部卒, 2003年同大学院情報系工学研究科博士後期課程修了, 博士(工学). 同年科学技術振興事業団戦略的創造研究推進事業(CREST)研究員. 同年神戸大学情報基盤センター助教. 2011年岡山県立大学情報工学部助教. 主にヒューマンインタラクション・アバタコミュニケーションの研究に従事. 2000年ヒューマンインタフェース学会学術奨励賞, 2002年・2004年ヒューマンインタフェース学会論文賞, 2003年情報処理学会大会奨励賞等受賞. 情報処理学会, 日本認知科学会, 日本人間工学会各会員.

### 高田 友寛 (学生会員)



2011年岡山県立大学情報工学部卒, 現在, 岡山県立大学大学院博士前期課程在学中. 主にヒューマンコミュニケーション, 身体的インタラクションの研究に従事.

### 渡辺 富夫 (正会員)



1983年東京大学大学院工学系研究科産業機械工学専攻博士課程修了(工学博士). 同年山形大学工学部情報工学科助手, 1984年同専任講師, 1989年同助教授. 1993年岡山県立大学情報工学部情報システム工学科教授. 1992年~1993年米国ブラウン大学ブラドリ・リサーチセンター客員研究員. ヒューマンインタラクション・コミュニケーションの研究に従事. 1998年・2003年・2012年IEEE RO-MAN, the Best Paper Award, 2001年・2002年・2004年・2005年ヒューマンインタフェース学会論文賞等受賞. 日本機械学会フェロー, 情報処理学会, 計測自動制御学会, 日本バーチャルリアリティ学会, 日本子ども学会, IEEE各会員.